

# Contribution of hidden modes to nonlinear epidemic dynamics in urban human proximity networks

Naoya Fujiwara,<sup>1,2,\*</sup> Abhijeet R. Sonawane,<sup>3,4,2</sup> Koji Iwayama,<sup>5,2</sup> and Kazuyuki Aihara<sup>2</sup>

<sup>1</sup>*Center for Spatial Information Science,*

*The University of Tokyo, 277-8568 Chiba, Japan*

<sup>2</sup>*Institute of Industrial Science, The University of Tokyo, 153-8505 Tokyo, Japan*

<sup>3</sup>*Channing Division of Network Medicine,*

*Brigham and Women's Hospital, Boston MA 02115, USA*

<sup>4</sup>*Harvard Medical School, Boston MA 02115, USA*

<sup>5</sup>*Faculty of Agriculture, Ryukoku University, Shiga 520-2194, Japan*

## Abstract

Recently developed techniques to acquire high-quality human mobility data allow large-scale simulations of the spread of infectious diseases with high spatial and temporal resolution. Analysis of such data has revealed the oversimplification of existing theoretical frameworks to infer the final epidemic size or influential nodes from the network topology. Here we propose a spectral decomposition-based framework for the quantitative analysis of epidemic processes on realistic networks of human proximity derived from urban mobility data. Common wisdom suggests that modes with larger eigenvalues contribute more to the epidemic dynamics. However, we show that hidden dominant structures, namely modes with smaller eigenvalues but a greater contribution to the epidemic dynamics, exist in the proximity network. This framework provides a basic understanding of the relationship between urban human motion and epidemic dynamics, and will contribute to strategic mitigation policy decisions.

Epidemics of infectious diseases in the human population, e.g. the SARS outbreak in 2002–2003 [1], 2009 H1N1 influenza pandemic [2], and Ebola outbreak [3], can be a serious factor in human mortality, and have a significant socio-economic impact in terms of reducing a population’s healthy years because of morbidity. In recent years, the mathematical modelling of the outbreak and spread of infectious diseases has mainly been performed using two approaches: agent-based models [4–10] and structured metapopulation models [11–22]. Both techniques incorporate real or synthetic data on the long-range mobility and migration of populations, but vary in their granularity of the population, from individuals to sub-groups of society. Various spatial scales have been studied, ranging from specific geographic locales such as cities [4, 5] to nations [6–8, 23, 24] and intra-/inter-continental regions [15, 16].

In particular, it is imperative to study the mechanisms of an epidemic’s spread in urban scenarios, considering restrictions on available space and transportation methods (e.g. traffic fluxes, travel routes, timescales of modes of transport), which result in high density and heterogeneous contact patterns among residents. These factors make the populace vulnerable to epidemics of diseases that spread through exhaled aerosols, e.g. [25]. To understand the details of epidemic dynamics within cities, it is crucial to exploit the heterogeneous contact patterns derived from human mobility data. Fortunately, recent technical advances in data acquisition methods enable us to incorporate human mobility data obtained from various sources [26–28] into mathematical models [16].

Network epidemiology [20, 29–33] is a useful means of uncovering the dependence of the contagion dynamics on the heterogeneous contact pattern. The approach of tracking connections between people provides a comprehensive framework with which to study the effects of the underlying network of people in the region of interest (e.g. cities). The present understanding of epidemic or social contagion processes has been bolstered by many network-based studies [16, 29, 34], which treat cities as social petri-dishes [35].

Analytical approaches such as heterogeneous mean field (HMF) theory [36–39] and moment closure approximation [40–42] provide an important bridge between the network topology and the spread of epidemics in complex networks. For example, the epidemic threshold is absent in scale-free networks with no degree correlation [36]. However, certain simplifications made to obtain analytical results, e.g. neglecting the degree correlation in the HMF [37], could be superfluous, and these approximations may not necessarily hold in realistic networks. Therefore, a novel analytical approach is required to better understand the spread

of epidemics in heterogeneous populations.

### Proximity networks and the SIR model

In this paper, we outline a framework for modelling a proximity network from transportation data, and propose an analytical method to estimate the infection probability of each individual from stochastic simulations. The prevalence, i.e. the average number of recovered agents who were once infected before the spread of the epidemic subsides, is of particular interest.

We can construct an individual-based network using a dataset of human mobility. The *People flow data* [28] record the one-day movement of individuals living in the Kanto region of Japan, which includes the Tokyo metropolitan area. Details of the dataset are given in the Methods section. We randomly chose the mobility track of  $N$  individuals from this dataset. Time-dependent proximity networks were then constructed by connecting individuals when they came within a certain geographical proximity, defined by a distance threshold  $D$  (Fig. 1). Let us denote the adjacency matrix of the proximity network as  $A(t)$ . All results in the main text were obtained with  $D = 1,000$  m and  $N = 10,000$ . Note that we observed the percolation-like transition and the scaling relation of the giant cluster component size ratio as  $S(N, D)/N \sim f(N^{0.605}D)$ . Using this scaling relation,  $D = 1,000$  m with  $N = 10,000$  corresponds to  $D \approx 8.3$  m with the real population in the corresponding region, where  $N \approx 27,500,000$ . A detailed discussion can be found in Supplementary Information SI1.

We simulated the disease propagation using the agent-based SIR model [43] on the above-mentioned network, which assigns susceptible (S), infected (I), or recovered (R) states to each agent. We assumed that each agent repeats the same trip pattern every day in the same way as recorded in the People flow data. When a susceptible agent is connected with an infected one, the former is infected with probability  $\beta\Delta t$  for a time interval  $\Delta t$ . If a susceptible agent is connected with  $d$  infected agents, they are infected with probability  $1 - (1 - \beta\Delta t)^d$  [29]. Further details of the stochastic simulation are given in the Methods section. In this study,  $\beta\Delta t$  is sufficiently small which enables us to perform the simulation with the discretised time step. An infected agent recovers with probability  $\mu\Delta t$  over this time interval. Once  $A(t)$ ,  $\beta$ , and  $\mu$  are given, stochastic simulations can be conducted to determine the epidemic dynamics. We denote the probabilities that agent  $j$  is in state S, I,

or R at time  $t$  as  $s_j(t)$ ,  $i_j(t)$ , and  $r_j(t)$ , respectively; the equality  $s_j(t) + i_j(t) + r_j(t) = 1$  holds. We define the infection probability of agent  $j$  as  $r_j(\infty)$ , because a recovered agent has experienced the infected state before recovering.

The epidemic dynamics can be characterized by three different stages [44]: the initial stage, in which stochastic fluctuations are dominant, the exponential growth stage, and the final stage, where nonlinearity suppresses the further spread of the disease. The latter two stages representing the dynamics of the outbreak are of particular interest. In these stages, the time evolution of the epidemic can be approximated by the following deterministic differential equations:

$$\begin{aligned}\frac{ds_j}{dt} &= -\beta s_j \sum_{k=1}^N \bar{A}_{jk} i_k, \\ \frac{di_j}{dt} &= \beta s_j \sum_{k=1}^N \bar{A}_{jk} i_k - \mu i_j, \\ \frac{dr_j}{dt} &= \mu i_j.\end{aligned}\tag{1}$$

Here, we approximate the time-dependent adjacency matrix  $A(t)$  with its time averaged form  $\bar{A}$  [45]. This averaging gives a sufficient approximation of the time evolution of  $s_j$ ,  $i_j$ , and  $r_j$  if the parameters  $\beta$  and  $\mu$  are sufficiently small, as discussed in Supplementary Information SI2. We also assume that node  $j$  being susceptible and node  $k$  being infected are statistically independent events [46]. Although this assumption does not hold for small networks [43], we have verified that the numerical solution of equation (25) gives a good approximation of the dynamics if  $D$  is much greater than the percolation transition point and the giant cluster component is sufficiently large.

One of the most common methods used to analyse the contagion processes within networks is HMF [36–39]. This approach assumes that, in a statistical sense, nodes grouped by the degrees behave in the same way. Under this assumption, one can derive the equation for the infection probability of each agent, and, in the absence of degree correlation, the epidemic threshold, i.e. the critical value of the epidemic’s spread, can be represented as a function of the first and second moments of the degree distribution [37]. One question that persists is whether such an approximation is valid in realistic networks of human contact. Actually, the accuracy of HMF depends on factors such as the mean degree and first-neighbour degree [47]. For the simplest case of the absence of degree correlation, where

HMF is often applied, the component of the first eigenvector is proportional to the degree [44]. However, this assumption does not hold for the proximity network (Fig. 2a). This highlights the need to develop an analytical framework to identify people with a high infection probability along with the overall infection propagation pattern. In this study, we have developed a method based on spectral decomposition and mode truncation which demonstrates that a small number of dominant modes, which may not necessarily have the largest eigenvalues, give a higher contribution in the prevalence of epidemics.

### Spectral analysis

In this section, we present a method for analysing the final epidemic size based on the spectral analysis of the averaged matrix  $\bar{A}$ , and elucidate the relevant dynamics of the epidemic's spread. The node-wise probabilities  $s_j(t)$ ,  $i_j(t)$ , and  $r_j(t)$  can be expanded as

$$s_j(t) = \sum_{a=1}^N \hat{s}_a(t) \phi_j^{(a)}, \quad i_j(t) = \sum_{a=1}^N \hat{i}_a(t) \phi_j^{(a)}, \quad r_j(t) = \sum_{a=1}^N \hat{r}_a(t) \phi_j^{(a)}, \quad (2)$$

where  $\phi_j^{(a)}$  denotes the  $a$ th eigenvector of  $\bar{A}$  associated with the eigenvalue  $\lambda_a$ . All eigenvalues of a real symmetric matrix  $\bar{A}$  are real, and can be labeled in descending order as  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N$ . The normalization condition  $\sum_{j=1}^N (\phi_j^{(a)})^2 = 1$  is adopted. Since  $\bar{A}$  is symmetric, the left and right eigenvectors coincide, and the expansion coefficients can be obtained with  $\hat{s}_a(t) = \sum_{j=1}^N s_j(t) \phi_j^{(a)}$ ,  $\hat{i}_a(t) = \sum_{j=1}^N i_j(t) \phi_j^{(a)}$ , and  $\hat{r}_a(t) = \sum_{j=1}^N r_j(t) \phi_j^{(a)}$ , respectively. Equation (25) can be rewritten in terms of these expansion coefficients (Supplementary Information SI3). The time evolution of the average number of infected and recovered agents,  $N_I(t) \equiv \sum_{j=1}^N i_j(t)$  and  $N_R(t) \equiv \sum_{j=1}^N r_j(t)$ , is plotted in Fig. 2b. The time series of  $i_j(t)$  is converted to that of  $\hat{i}_a(t)$  as in Fig. 2c.

In the exponential stage of the epidemic spreading,  $1 - s_j$ ,  $i_j$ , and  $r_j$  are small. The exponential growth of each coefficient  $\hat{i}_a(t)$  can be described by the following linearised equation [48]

$$\frac{d\hat{i}_a}{dt} \approx (\beta\lambda_a - \mu)\hat{i}_a \quad (3)$$

(see Supplementary information SI4 for the derivation of this expression from the nonlinear mode coupling equation). Hence,  $\hat{i}_a(t) = \hat{i}_a(0)e^{(\beta\lambda_a - \mu)t}$  holds. It is evident from this equation that a mode with a larger eigenvalue  $\lambda_a$  grows faster in this stage. In particular,

the exponential growth rate of the number of infected agents is characterised by  $\beta\lambda_1 - \mu$ . The stochastic simulation of the agent-based model verifies that  $\hat{i}_1$  is indeed dominant in the initial stage (Fig. 2c). In addition, the epidemic threshold is related to  $\lambda_1$  as  $\beta_c = \mu/\lambda_1$  (Fig. 2d) [48].

However, as shown in Fig. 2c, the coefficients  $\hat{i}_a(t)\langle\phi^{(a)}\rangle$  ( $a = 2, 3, \dots, N$ ) are comparable to  $\hat{i}_1(t)\langle\phi^{(1)}\rangle$  in the later stages of the epidemic, where  $\langle v \rangle = \sum_{j=1}^N v_j/N$  is the node-wise average of a vector  $v$ . The importance of each mode needs to be assessed to ascertain its dominance in the spreading mechanism. This can be achieved by introducing a quantifier of the *contribution* of each mode to the prevalence as

$$C_a \equiv \hat{r}_a(\infty)\langle\phi^{(a)}\rangle. \quad (4)$$

The definition of this contribution enables us to account for the prevalence as the sum of the contributions of all modes,  $\sum_{j=1}^N r_j(\infty) = \sum_{a=1}^N C_a$ . It is worth introducing the contribution  $C_a^{\text{all}}$  for the case where all agents are infected, i.e.  $r_j(\infty) = 1$  for all  $j$ . Since the coefficients satisfy  $\hat{r}_a(\infty) = N\langle\phi^{(a)}\rangle$  in this case, one can easily verify that  $C_a^{\text{all}} = N\langle\phi^{(a)}\rangle^2$  holds. Interestingly, the contribution  $C_a^{\text{all}}$  does not monotonically increase with the eigenvalue, although a positive correlation can be observed (Fig. 3a). Since  $C_a$  and  $C_a^{\text{all}}$  are functions of the average of the eigenvector components, a mode associated with a smaller eigenvalue may contribute more than one with a larger eigenvalue. Although the larger eigenvalue denotes faster growth in the exponential regime, it does not necessarily mean a larger contribution to the prevalence.

It is interesting to note that the “hidden” important structures, corresponding to modes that make a large contribution to the prevalence but have smaller eigenvalues, are unravelled through the spectral analysis. It is still possible to detect such modes when the prevalence is much smaller than the all-infected case (Fig. 3b). The heterogeneity in the contribution is evident from the cumulative contribution  $\Gamma_n \equiv \sum_{a=1}^n C_{\sigma(a)}$  and  $\Gamma_n^{\text{all}} \equiv \sum_{a=1}^n C_{\sigma(a)}^{\text{all}}$  (Fig. 3c). Here, the modes are sorted such that  $C_{\sigma(1)}^{\text{all}} \geq C_{\sigma(2)}^{\text{all}} \geq \dots \geq C_{\sigma(N)}^{\text{all}}$ , where  $\sigma(a)$  denotes the permutation of the index set  $\{1, 2, 3, \dots, N\}$ . Note that  $\Gamma_N^{\text{all}} = 1$  holds, because the infection pattern is represented using all modes. Figure 3c suggests  $\Gamma_{1000}^{\text{all}} \approx 0.9$  for the all-infected case, indicating that approximately 90% of the prevalence is described by the top 10% of modes for the present proximity network. This figure indicates that the number of modes needed to describe the prevalence decreases if the prevalence is smaller than 1.

The above observations led us to the idea of describing the epidemic dynamics with a small number of dominant modes that give a high contribution. In the case of the agent-based SIR dynamics, we can derive the truncated equation for the contributions at the final time with  $M$  ( $\leq N$ ) modes,  $\tilde{C}_{\sigma(a)}^{(M)}(\infty)$ , as

$$\begin{aligned} \tilde{C}_{\sigma(a)}^{(M)}(\infty) = N \langle \phi^{(\sigma(a))} \rangle^2 - \sum_{j=1}^N s_j(0) \phi_j^{(\sigma(a))} \langle \phi^{(\sigma(a))} \rangle \\ \times \exp \left\{ -\frac{\beta}{\mu} \sum_{b=1}^M \lambda_{\sigma(b)} [\tilde{C}_{\sigma(b)}^{(M)}(\infty) - \tilde{C}_{\sigma(b)}^{(M)}(0)] \frac{\phi_j^{(\sigma(b))}}{\langle \phi^{(\sigma(b))} \rangle} \right\}, \quad (5) \end{aligned}$$

where  $a = 1, 2, \dots, M$ . This equation is derived in the Methods section. This transcendental equation approximates the epidemic dynamics with fewer effective degrees of freedom. If we use all the modes, i.e.  $M = N$ , the solution of equation (5) coincides with the solutions of equation (25) for  $t \rightarrow \infty$ . Furthermore, the solution of this equation computed with Newton's method provides a good description of the contribution obtained by the stochastic simulation (Fig. 3d). We further note that considering fewer than  $N$  modes suffices to approximate the prevalence as obtained by the stochastic simulation (Figs. 2d, 3c). Hence, this procedure replaces the need to solve an ordinary differential equation in  $2N$  independent variables (25) with the requirement to use information from the  $M$  dominant modes.

As discussed above, the essence of theoretical analyses of epidemic spreading within networks is to infer the dynamical information from the network topology. Conventionally, HMF [36–39] uses degree distributions and degree correlations as topological information. In contrast, our approach uses spectral information, which can take more general properties of the network topology into account. As described in the Methods, the final size equation in HMF [37] is derived from equation (5) under the following simple assumptions: (i) degree correlation is absent, and (ii) the first mode almost perfectly governs the dynamics. As shown in Fig. 2a, the first assumption is not satisfied in the proximity network. Moreover, it is evident from Figs. 2d and 3a that considering only the first mode is not sufficient to account for the whole epidemic dynamics, as  $C_1^{\text{all}}$  is less than 6%. The contribution of other dominant modes must be taken into account to quantify the epidemic dynamics (Fig. 3d). Thus, HMF with no degree correlation is not a sufficient approximation for the proximity network. It would be possible to improve the approximation by considering the degree correlation in HMF [38], but it is generally difficult to include information about the

second-nearest neighbours. In contrast, the proposed approach provides the flexibility to improve the approximation by using an arbitrary number of dominant modes.

### Identification of the spatio-temporal hot spots of infection

Another feature of epidemic spreading in proximity networks is the strong heterogeneity of the infection risk in space and time. It is important to identify such locations that have a high probability of infection, namely “hot spots”, for the mitigation and management of the epidemic. The spatial distribution of the hot spots can be mapped to provide an insight into the nature of epidemic spreading in urban areas (Fig. 4). More importantly, to uncover the relevant epidemic dynamics, such identification should be conducted with fewer parameters than the number of agents.

For the present model to be predictive, let us infer the infection probability of a test agent from the mobility pattern of other agents. The probability of infection of this agent depends on the presence of other infectious agents in the vicinity, which in turn varies from region to region and at different times of day,  $0 \leq \tau \leq T = 1,440$  min. Let  $\mathbf{x} = (x, y)^T$  be a position vector and  $\mathbf{x}_j(t)$  be the position of agent  $j$  at time  $t$ . We introduce a kernel function  $K(\mathbf{x})$  signifying the presence or absence of any interaction between two agents based on the distance between them. Since agents are connected within the threshold  $D$ , the Heaviside step function  $K(\mathbf{x}) = \Theta(|\mathbf{x}| - D)$  is adopted as the kernel function. This allows us to define a spatio-temporal risk factor of infection  $\rho(\tau, \mathbf{x})$  as

$$\rho(\tau, \mathbf{x}) \equiv \sum_{j=1}^N r_j(\infty) K(\mathbf{x} - \mathbf{x}_j(\tau)), \quad (6)$$

which can be interpreted as the weighted sum of probabilities  $r_j(\infty)$ . Integrating  $\rho(\tau, \mathbf{x}_p(\tau))$  along the trajectory of the test agent  $\mathbf{x}_p(\tau)$ , one can obtain the infection probability of this agent. Details of the derivation are discussed in Supplementary Information SI5. Figure 4a depicts the spatial distribution of the risk factor  $\rho(\tau, \mathbf{x})$  across the Kanto region of Japan. In the present model, which assumes homogeneous parameter values of  $\beta$  and  $\mu$ , the high-risk area during the daytime is the central business district, where many workers are located. In the nighttime, certain suburban areas from where people commute to the centre become high risk. Furthermore, people travelling to the central business district from their suburban residences allow the disease to spread over the whole metropolitan area. Figure 4c shows



that the infection probability of the test agents can be estimated from the integration of  $\rho(\tau, \mathbf{x}_p(\tau))$ , obtained from stochastic simulation, along its trajectory.

Instead of counting the infection probability  $r_j(\infty)$  in the neighbourhood of the region of interest, the spectral properties of the averaged adjacency matrix can be incorporated into the risk factor by choosing the first few high-contribution modes. Namely, one can approximate  $r_j(\infty)$  and  $\rho(\tau, \mathbf{x})$  as  $\tilde{r}_j^{(M)}(\infty) = \sum_{a=1}^M \hat{r}_{\sigma(a)}(\infty) \phi_j^{(\sigma(a))}$  and  $\tilde{\rho}^{(M)}(\tau, \mathbf{x}) = \sum_{j=1}^N \tilde{r}_j^{(M)}(\infty) K(\mathbf{x} - \mathbf{x}_j(\tau))$  using  $M$  modes. Hazard maps based on the mode truncation at different times are similar to those given by the stochastic simulation (Fig. 4a). Moreover, Fig. 4d shows that the integration of  $\tilde{\rho}^{(100)}(\tau, \mathbf{x}_p(\tau))$  along the path of the test agents predicts their infection probability. Thus, we conclude that mode truncation enables us to describe the spatio-temporal infection risk with a small number of modes. In other words, few modes that have high contributions dominate the spread of the epidemic.

## Conclusion

Data on human mobility, particularly contact information of urban residents, play a key role in the analysis of an epidemic spreading. Such high-resolution data allow detailed individual-level modelling, rather than coarse-grained metapopulation studies. At the same time, novel theoretical tools can be used to uncover the dominant dynamics at work in data-driven models without the need for oversimplification. Linear analysis characterises the exponential growth in the initial stage; nonetheless, this study clarified that the final size of the epidemic is provided by the analysis of various modes. The mode truncation described here revealed that the prevalence of epidemics is actually dominated by a small number of modes as compared to the number of agents, i.e. the effective number of degrees of freedom is much smaller than the system size. The proposed method practically limits the expanse of mitigation policies such as targeted vaccination, and provides specific alerts by following the spreading trail of infection. This also quantifies the spatio-temporal risk associated with empirical travel patterns. The analytical framework and realistic mobility data presented here provide a promising starting point for the study and possible confrontation of epidemics in the real world. Future studies should examine ways to facilitate the rapid decay of dominant modes using dynamic intervention strategies.

## Methods

### People flow data and construction of the proximity network

People flow data were generated based on person–trip surveys made by Japan’s Ministry of Land, Infrastructure, Transport, and Tourism. The surveys were based on questionnaires that asked about basic individual information such as gender, age, occupation, places visited, and trip modes on one day. For the sake of privacy, spatial information was recorded with a resolution such that the complete location could not be determined. The questionnaires allowed us to access advanced information such as transportation modes, travel times, transfer points, and trip purposes. Traces of the trips between two places reported in the raw data were interpolated using geographical information systems methods [28].

The proximity network was constructed by placing a link between two individuals at time  $t$  if they were within the threshold distance  $D$  of one another (Fig. 1b). We added another constraint by only placing a link between individuals using the same transportation mode; it is impractical to connect individuals walking along the road with those traveling in an automobile or train because of the improbability of the transmission of infection. The connectivity of the network can be described by the time-dependent adjacency matrix  $A(t)$ , where  $A_{jk}(t) = A_{kj}(t) = 1$  if individuals  $j$  and  $k$  are connected at time  $t$  and  $A_{jk}(t) = A_{kj}(t) = 0$  otherwise. This allowed us to obtain networks for  $N$  agents at every  $\Delta t$  time interval. Proximity networks are spatially embedded temporal networks, which appear in various forms of transportation and infrastructure networks [49].

### Settings of the stochastic simulations

The agent-based SIR model simulation was performed with the same time-dependent adjacency matrix  $A(t)$ . The time evolution of the probabilities  $s_j(t)$ ,  $i_j(t)$ , and  $r_j(t)$  was computed using 500 independent stochastic simulation runs with different random initial conditions. We also assumed that the human mobility patterns had a periodicity of 24 h owing to the circadian nature and economic situation of people going to their place of work and back home. Parameter values were taken as  $\Delta t = 10$  min,  $\mu = 2.0 \times 10^{-4} \text{ min}^{-1}$ , and  $\beta = 1.5 \times 10^{-5} \text{ min}^{-1}$ , except for Fig. 2d, in which  $\beta$  varies.

### Derivation of equation (5)

One of the main results of this article, the equation for the asymptotic solution of equation (5), is derived as follows. Substituting the expansion (26) into equation (25), we obtain

$$\frac{ds_j}{dt} = -\beta s_j \sum_{a=1}^N \lambda_a \hat{i}_a(t) \phi_j^{(a)}. \quad (7)$$

Let us define

$$\Psi_j(t) \equiv \sum_{a=1}^N \int_0^t \lambda_a \hat{i}_a(t') \phi_j^{(a)} dt' \quad (8)$$

$$= \frac{1}{\mu} \sum_{a=1}^N \lambda_a [\hat{r}_a(t) - \hat{r}_a(0)] \phi_j^{(a)}, \quad (9)$$

so that the solution of this equation is given as

$$s_j(t) = s_j(0) e^{-\beta \Psi_j(t)}, \quad (10)$$

or equivalently,

$$\hat{s}_a(t) = \sum_{j=1}^N s_j(0) \phi_j^{(a)} e^{-\beta \Psi_j(t)}. \quad (11)$$

For  $t \rightarrow \infty$ ,  $i_j(\infty)$  should vanish, and the conservation of probability condition is given as  $s_j(\infty) + r_j(\infty) = 1$ . This is equivalent to

$$\hat{r}_a(\infty) = N \langle \phi^{(a)} \rangle - \hat{s}_a(\infty). \quad (12)$$

Substituting equation (11) into this expression, we obtain

$$\hat{r}_a(\infty) = N \langle \phi^{(a)} \rangle - \sum_{j=1}^N s_j(0) \phi_j^{(a)} \exp \left\{ -\frac{\beta}{\mu} \sum_{b=1}^N \lambda_b [\hat{r}_b(\infty) - \hat{r}_b(0)] \phi_j^{(b)} \right\}. \quad (13)$$

This is a transcendental equation for  $\hat{r}_a(\infty)$ , and one can determine the infection probability by solving this equation instead of solving the differential equation. We choose  $M$  ( $\leq N$ ) modes and neglect the other modes to obtain

$$\hat{r}_{\sigma(a)}(\infty) \approx N \langle \phi^{(\sigma(a))} \rangle - \sum_{j=1}^N s_j(0) \phi_j^{(\sigma(a))} \exp \left\{ -\frac{\beta}{\mu} \sum_{b=1}^M \lambda_{\sigma(b)} [\hat{r}_{\sigma(b)}(\infty) - \hat{r}_{\sigma(b)}(0)] \phi_j^{(\sigma(b))} \right\}, \quad (14)$$

where  $\sigma(a)$  and  $\sigma(b)$  denote arbitrary permutations of modes. Finally, we multiply both sides by  $\langle \phi^{(\sigma(a))} \rangle$ , and obtain equation (5) by choosing the top  $M$  modes with respect to their contribution.

## Derivation of the final size equation in heterogeneous mean field theory

The relationship between the equation for the expansion coefficients (13) and that obtained with HMF can be described as follows. We concentrate on the simplest case, where the probability  $P_{dd'}$  that a node with degree  $d$  and one with degree  $d'$  are connected is proportional to the product of their degrees, i.e.  $P_{dd'} = dd'P(d')/\langle d \rangle$ , where  $P(d)$  denotes the degree distribution of the network satisfying  $\sum_{d=1}^{N-1} P(d) = 1$ . In this case, the largest eigenvalue of the adjacency matrix is  $\lambda_1 = \langle d^2 \rangle / \langle d \rangle$ , and the component of the corresponding normalised eigenvector is  $\phi_j^{(1)} = d_j / \sqrt{N \langle d^2 \rangle}$ . Here, we denote the degree of node  $j$  as  $d_j$ . We assume that the probability of nodes with the same degree is identical, i.e.  $r_j(t) = \bar{r}_{d_j}(t)$  holds.

Substituting the expression for the first eigenvector into equation (13) and neglecting other modes with the initial condition  $s_j(0) \approx 1$ , we get

$$\hat{r}_1(\infty) \approx N \langle \phi^{(1)} \rangle - \sum_{j=1}^N \phi_j^{(1)} e^{-\frac{\beta}{\mu} \lambda_1 \hat{r}_1(\infty) \phi_j^{(1)}} \quad (15)$$

$$= \sqrt{\frac{N}{\langle d^2 \rangle}} \left[ \langle d \rangle - \sum_{d=1}^{N-1} d P(d) e^{-\frac{\beta}{\mu} \langle d^2 \rangle \hat{r}_1(\infty) d / (\langle d \rangle \sqrt{N \langle d^2 \rangle})} \right], \quad (16)$$

where the sum in the second equation is taken for degree  $d$ . Note that

$$\hat{r}_1(t) = \sum_{j=1}^N r_j(t) \phi_j^{(1)} = \sqrt{\frac{N}{\langle d^2 \rangle}} \sum_{d=1}^{N-1} d P(d) \bar{r}_d(t) \quad (17)$$

$$= \sqrt{\frac{N}{\langle d^2 \rangle}} \mu \langle d \rangle \Psi(t) \quad (18)$$

holds, where we have defined

$$\Psi(t) \equiv \frac{1}{\mu \langle d \rangle} \sum_{d=1}^{N-1} d P(d) \bar{r}_d(t). \quad (19)$$

Substituting this equation into equation (16) and taking the limit  $t \rightarrow \infty$ , we obtain

$$\mu \Psi(\infty) = 1 - \frac{1}{\langle d \rangle} \sum_{d=1}^{N-1} d P(d) e^{-\beta d \Psi(\infty)}. \quad (20)$$

This is the equation for  $\Psi(\infty)$  in HMF derived in [37].

---

\* Electronic address: fujiwara@csis.u-tokyo.ac.jp

- [1] Peiris, J. S. M. *et al.* Clinical progression and viral load in a community outbreak of coronavirus-associated SARS pneumonia: a prospective study. *Lancet* **361**, 1767–1772 (2003).
- [2] Fraser, C. *et al.* Pandemic potential of a strain of influenza A (H1N1): early findings. *Science* **324**, 1557–1561 (2009).
- [3] Gire, S. K. *et al.* Genomic surveillance elucidates ebola virus origin and transmission during the 2014 outbreak. *Science* **345**, 1369–1372 (2014).
- [4] Eubank, S. *et al.* Modelling disease outbreaks in realistic urban social networks. *Nature* **429**, 180–184 (2004).
- [5] Saito, M. M. *et al.* Enhancement of collective immunity in Tokyo metropolitan area by selective vaccination against an emerging influenza pandemic. *PLoS ONE* **8**, e72866 (2013).
- [6] Ferguson, N. M. *et al.* Strategies for containing an emerging influenza pandemic in southeast Asia. *Nature* **437**, 209–214 (2005).
- [7] Ajelli, M. & Merler, S. An individual-based model of hepatitis A transmission. *J. Theor. Biol.* **259**, 478–488 (2009).
- [8] Merler, S. & Ajelli, M. The role of population heterogeneity and human mobility in the spread of pandemic influenza. *Proc. Biol. Sci.* **277**, 557–565 (2010).
- [9] Ferguson, N. M. *et al.* Strategies for mitigating an influenza pandemic. *Nature* **442**, 448–452 (2006).
- [10] Halloran, M. E. *et al.* Modeling targeted layered containment of an influenza pandemic in the United States. *Proc. Natl. Acad. Sci.* **105**, 4639–4644 (2008).
- [11] Wang, B., Cao, L., Suzuki, H. & Aihara, K. Epidemic spread in adaptive networks with multitype agents. *J. Phys. A* **44**, 035101 (2011).
- [12] Colizza, V., Pastor-Satorras, R., & Vespignani, A. Reaction–diffusion processes and metapopulation models in heterogeneous networks. *Nature Phys.*, **3**, 276–282 (2007).
- [13] Balcan, D. & Vespignani, A. Phase transitions in contagion processes mediated by recurrent mobility patterns. *Nature Phys.* **7**, 581–586 (2011).
- [14] Poletto, C., Tizzoni, M., & Colizza, V. Heterogeneous length of stay of hosts’ movements and spatial epidemic spread. *Sci. Rep.* **2**, 476 (2012).
- [15] Colizza, V., Barrat, A., Barthélemy, M. & Vespignani, A. The role of the airline transportation network in the prediction and predictability of global epidemics. *Proc. Natl. Acad. Sci.* **103**, 2015–2020 (2006).

- [16] Colizza, V., Barrat, A., Barthélemy, M., Valleron, A.-J. & Vespignani, A. Modeling the worldwide spread of pandemic influenza: Baseline case and containment interventions. *PLoS Med.* **4**, e13 (2007).
- [17] Balcan, D. *et al.* Multiscale mobility networks and the spatial spreading of infectious diseases. *Proc. Natl. Acad. Sci.* **106**, 21484–21489 (2009).
- [18] Sattenspiel, L. & Dietz, K. A structured epidemic model incorporating geographic mobility among regions. *Math. Biosci.* **128**, 71–91 (1995).
- [19] Vazquez, A. Epidemic outbreaks on structured populations. *J. Theor. Biol.* **245**, 125–129 (2007).
- [20] Watts, D. J., Muhamad, R., Medina, D. C. & Dodds, P. S. Multiscale, resurgent epidemics in a hierarchical metapopulation model. *Proc. Natl. Acad. Sci.* **102**, 11157–11162 (2005).
- [21] Rohani, P., Earn, D. J. D. & Grenfell, B. T. Opposite patterns of synchrony in sympatric disease metapopulations. *Science* **286**, 968–971 (1999).
- [22] Rvachev, L. A. & Longini Jr, I. M. A mathematical model for the global spread of influenza. *Math. Biosci.* **75**, 3–22 (1985).
- [23] Riley, S. Large-scale spatial-transmission models of infectious disease. *Science* **316**, 1298–1301 (2007).
- [24] Grais, R. F., Ellis, J. H. & Glass, G. E. Assessing the impact of airline travel on the geographic spread of pandemic influenza. *Eur. J. Epidemiol.* **18**, 1065–1072 (2003).
- [25] Tellier, R. Aerosol transmission of influenza A virus: a review of new studies. *J. R. Soc. Interface* **6 Suppl 6**, S783–S790 (2009).
- [26] Onnela, J. P. *et al.* Structure and tie strengths in mobile communication networks. *Proc. Natl. Acad. Sci.* **104**, 7332–7336 (2007).
- [27] Cattuto, C. *et al.* Dynamics of person-to-person interactions from distributed RFID sensor networks. *PLoS ONE* **5**, e11596 (2010).
- [28] Sekimoto, Y., Shibasaki, R., Kanasugi, H., Usui, T. & Shimazaki, Y. Pflow: Reconstructing people flow recycling large-scale social survey data. *Pervasive Comput.* **10**, 27–35 (2011).
- [29] Masuda, N. & Holme, P. Predicting and controlling infectious disease epidemics using temporal networks. *F1000Prime Rep.* **5**, 6 (2013).
- [30] Danon, L. *et al.* Networks and the epidemiology of infectious disease. *Interdiscip. Perspect. Infect. Dis.* **2011**, 284909 (2011).

- [31] Pastor-Satorras, R., Castellano, C., Van Mieghem, P. & Vespignani, A. Epidemic processes in complex networks. *Rev. Mod. Phys.* **87**, 925–979 (2015).
- [32] Kitsak, M. *et al.* Identification of influential spreaders in complex networks. *Nature Phys.* **6**, 888–893 (2010).
- [33] Brockmann, D. & Helbing, D. The hidden geometry of complex, network-driven contagion phenomena. *Science* **342**, 1337–1342 (2013).
- [34] Salathé, M. *et al.* A high-resolution human contact network for infectious disease transmission. *Proc. Natl. Acad. Sci.* **107**, 22020–22025 (2010).
- [35] Kitchovitch, S. & Liò, P. Risk perception and disease spread on social networks. *Procedia Comput. Sci.* **1**, 2345–2354 (2010).
- [36] Pastor-Satorras, R. & Vespignani, A. Epidemic spreading in scale-free networks. *Phys. Rev. Lett.* **86**, 3200–3203 (2001).
- [37] Moreno, Y., Pastor-Satorras, R. & Vespignani, A. Epidemic outbreaks in complex heterogeneous networks. *Eur. Phys. J. B* **26**, 521–529 (2002).
- [38] Boguná, M., Pastor-Satorras, R. & Vespignani, A. Epidemic spreading in complex networks with degree correlations. In *Statistical Mechanics of Complex Networks Notes in Physics*, Springer, Berlin, 2003.
- [39] Yang, R. *et al.* Epidemic spreading on heterogeneous networks with identical infectivity. *Phys. Lett. A* **364**, 189–193 (2007).
- [40] Keeling, M. J. The effects of local spatial structure on epidemiological invasions. *Proc. R. Soc. B* **266**, 859–867 (1999).
- [41] Eames, K. T. D. & Keeling, M. J. Modeling dynamic and network heterogeneities in the spread of sexually transmitted diseases. *Proc. Natl. Acad. Sci.* **99**, 13330–13335 (2002).
- [42] Boots, M. & Sasaki, A. Parasite-driven extinction in spatially explicit host-parasite systems. *Am. Nat.* **159**, 706–713 (2002).
- [43] Youssef, M. & Scoglio, C. An individual-based approach to SIR epidemics in contact networks. *J. Theor. Biol.* **283**, 136 – 144 (2011).
- [44] Barrat, A., Barthélemy, M. & Vespignani, A. *Dynamical processes on complex networks* (Cambridge Univ. Press, 2008).
- [45] Stehlé, J., *et al.* Simulation of an SEIR infectious disease model on the dynamic contact network of conference attendees. *BMC Med.* **9**, 87 (2011).

- [46] Newman, M. *Networks: An introduction* (Oxford Univ. Press, 2010).
- [47] Gleeson, J. P., Melnik, S., Ward, J. A., Porter, M. A. & Mucha, P. J. Accuracy of mean-field theory for dynamics on real-world networks. *Phys. Rev. E* **85**, 026106 (2012).
- [48] Wang, Y., Chakrabarti, D., Wang, C. & Faloutsos, C. Epidemic spreading in real networks: An eigenvalue viewpoint. In *Proc. 22nd International Symposium on Reliable Distributed Systems* 25–34 (2003).
- [49] Barthélemy, M. Spatial networks. *Phys. Rep.* **499**, 1 – 101 (2011).

### **Acknowledgement**

N.F. is grateful for the stimulating discussions with T. Takaguchi, T. Aoki, Y. Sughiyama, and Y. Yazaki. This research is supported by the Aihara Project, the FIRST program from JSPS, initiated by CSTP, and by CREST, JST. This research is the result of the joint research with Center for Spatial Information Science, the University of Tokyo (No. 315). N.F. is supported by JSPS KAKENHI Grant Number 15K16061.

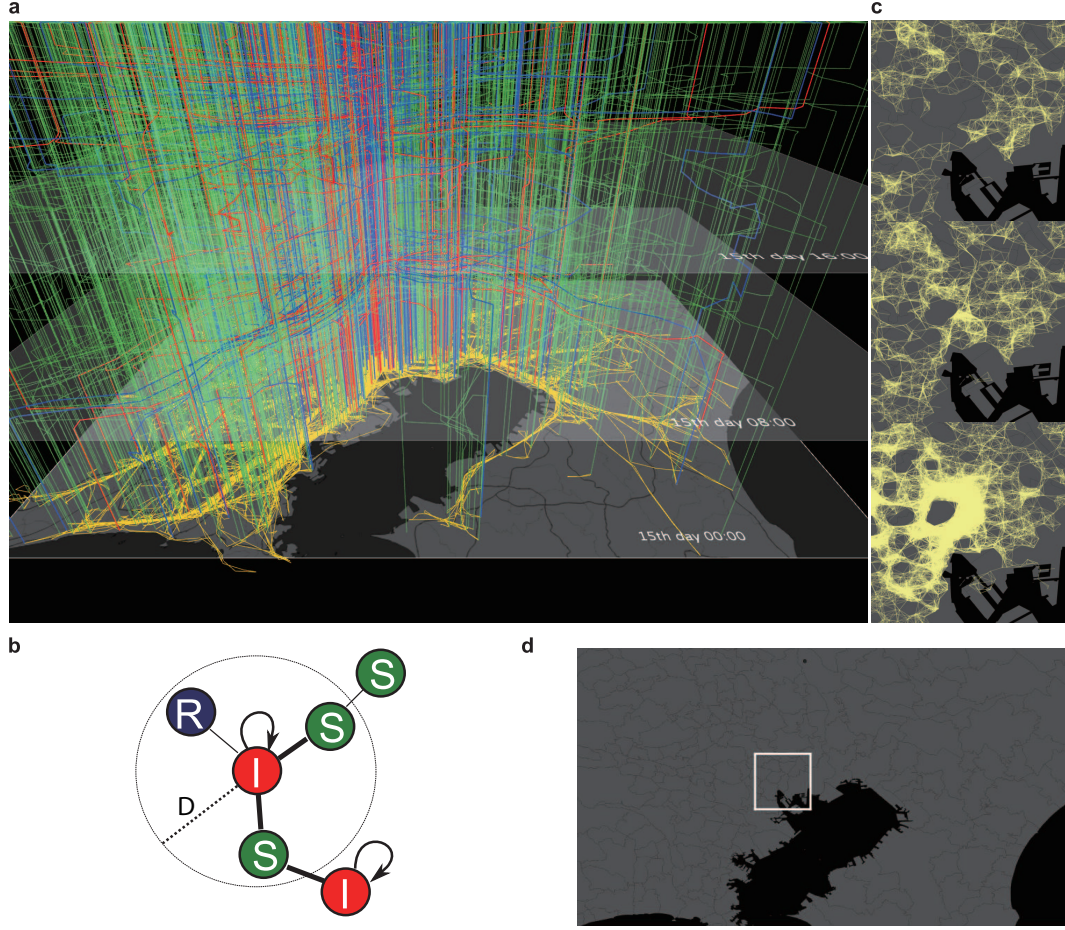
### **Author contributions**

All authors designed the research. N.F. and A.R.S. carried out the numerical simulations. N.F., A.R.S, and K.I. analysed the results. N.F., K.I., and K.A. contributed to the analytical calculations. N.F. and A.R.S. wrote the bulk of the manuscript. K.I. and K.A. edited the manuscript.

### **Competing financial interests**

The authors declare no competing financial interests.





**FIG. 1: Illustration of epidemic spreading in cities and the time-dependent proximity network.** a. Mobility pattern and infection state of agents in the Tokyo metropolitan area. Colours represent susceptible (green), infected (red), and recovered (blue) states on the 15th day of an agent-based stochastic simulation run. The vertical axis represents time. On the background map, dark grey and black regions represent land and bay areas, respectively. Black lines in the land are railways and highways. Yellow lines are the projection of the trajectories. Density of the yellow lines indicates the traffic density in the region. The centre of the map, which has the highest yellow line density, corresponds to the central business district (CBD) of Tokyo. As seen from the figure, most infected and recovered agents move to the CBD in the morning and return to their hometowns in the evening. b. Schematic figure of construction of the proximity network from the People flow data. Infection takes place between susceptible (S) and infected (I) agents within a distance  $D$  with probability  $\beta\Delta t$  for a time interval  $\Delta t$ . Recovered (R) agents do not affect infection of other agents. c. The connectivity pattern between agents near the CBD for  $D = 1,000$  m and  $N = 10,000$  at 00:00, 08:00, and 12:00. The region of the map is indicated by the white square in the wide-area map in panel d.

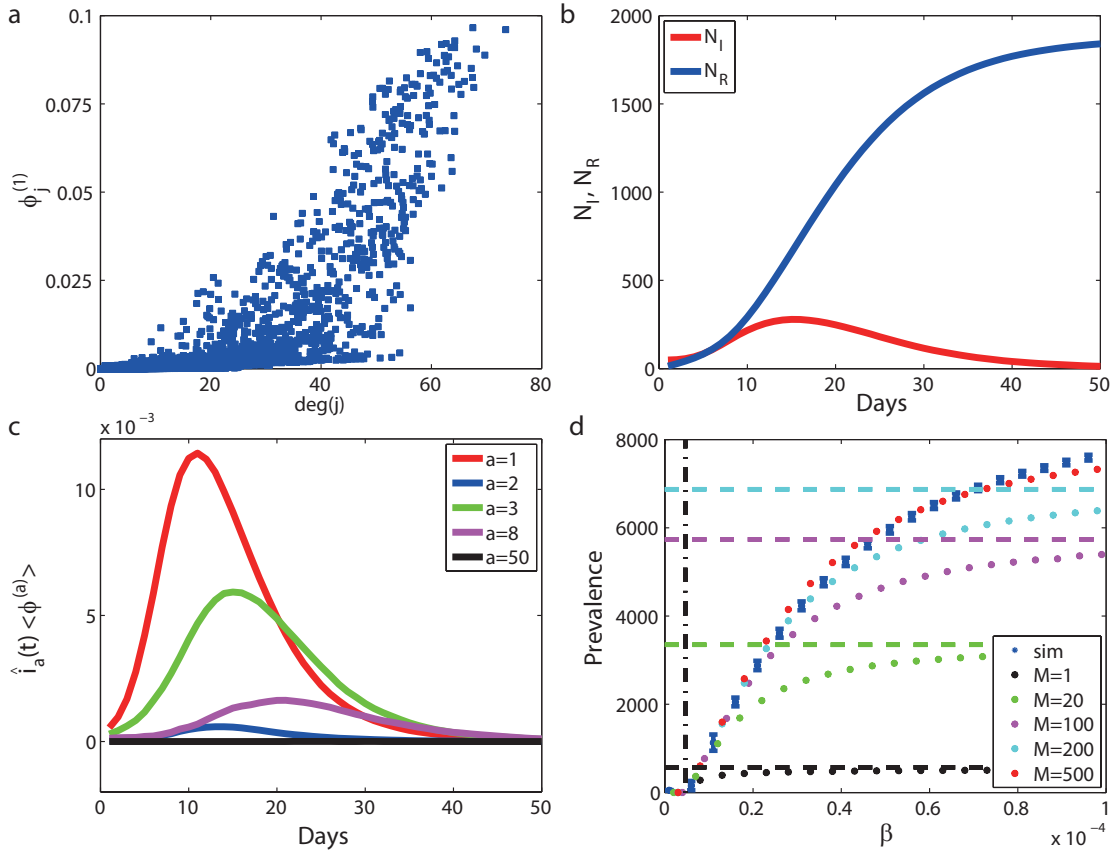
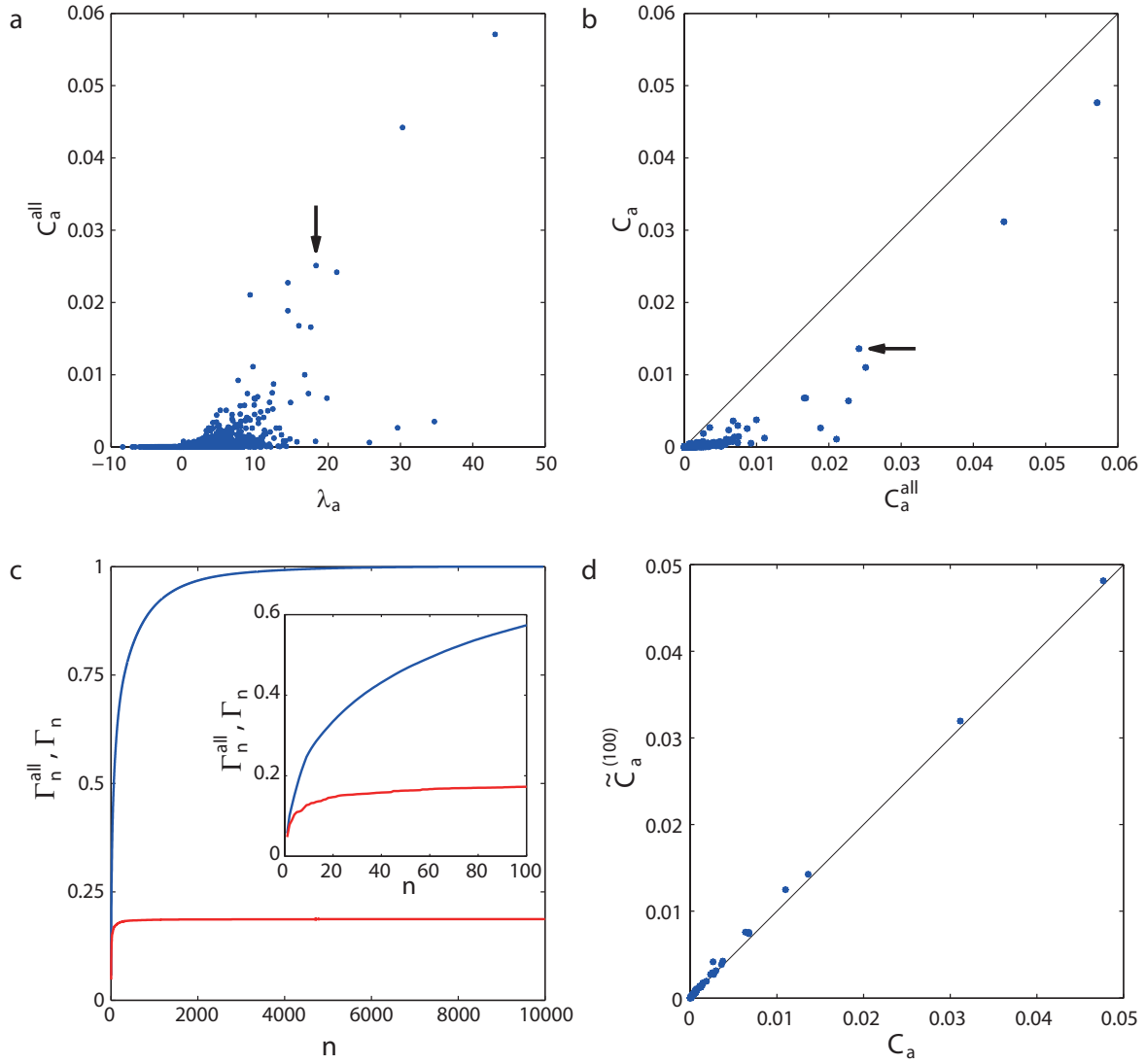
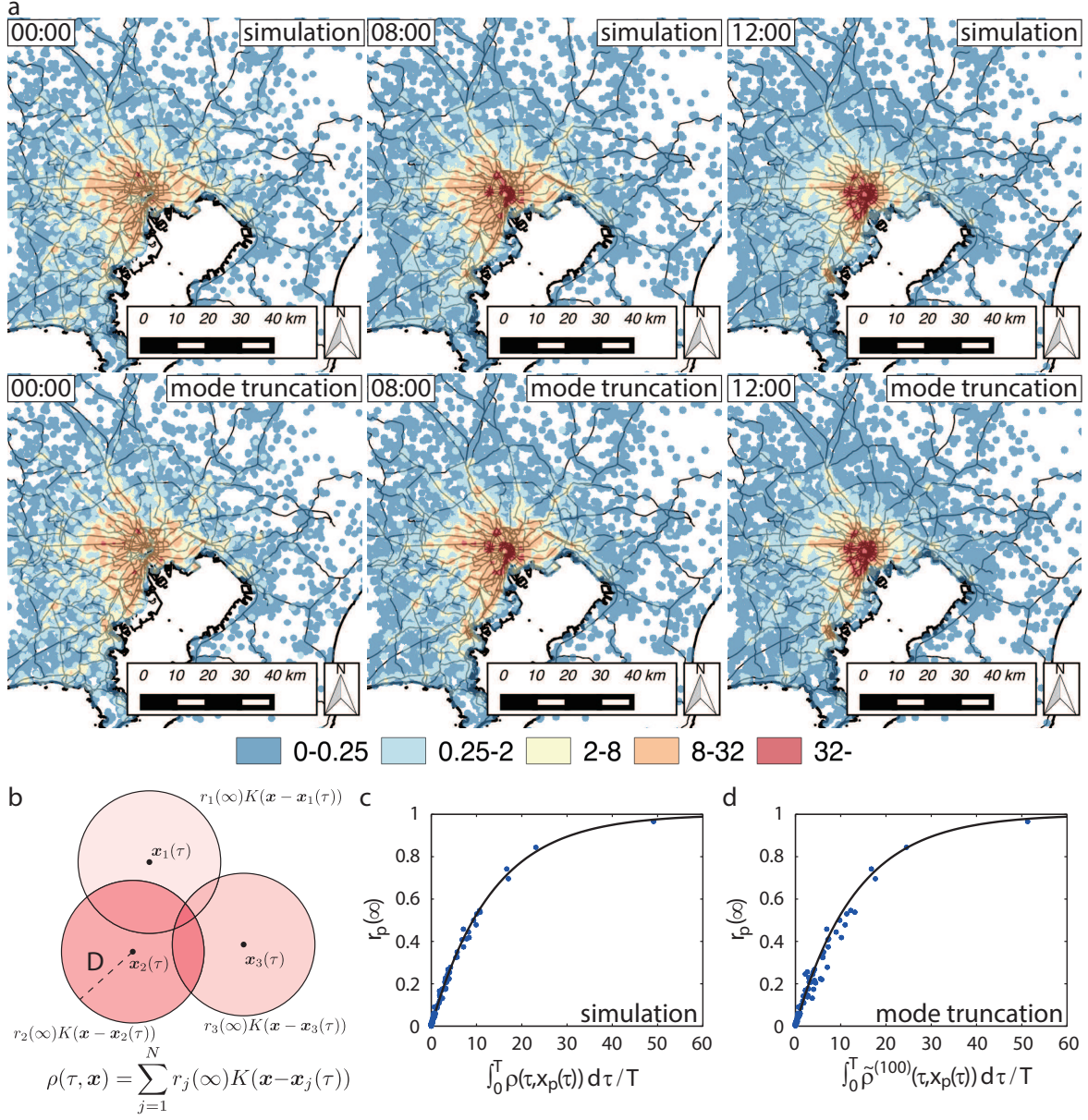


FIG. 2: **Spectral method for the analysis of epidemic spreading.** a. The first eigenvector component versus the degree of  $\bar{A}$ . For networks with no degree correlation, the component of the first eigenvector is proportional to the degree, but this result shows that this is not the case for the proximity network. b. Time evolution of the average numbers of infected  $N_I$  and recovered  $N_R$  people in the stochastic simulation. The number of infected agents increases exponentially, and then decays because of the nonlinear effect. c. Time evolution of  $\hat{i}_a(t) \langle \phi^{(a)} \rangle$ . Initially, all coefficients grow exponentially, and the mode with the largest eigenvalue  $\hat{i}_1(t) \langle \phi^{(1)} \rangle$  grows fastest. In the later stages of infection, other modes  $\hat{i}_a(t)$  ( $a \neq 1$ ) grow and  $\hat{i}_1(t)$  decays faster than the others. Hence, the other modes are not negligible. However, modes with small eigenvalues are negligible because their amplitude is sufficiently small. Therefore, much smaller than  $N$  (but not one) modes should be taken into account. d. Dependence of the prevalence on the infection rate  $\beta$ . Results of the stochastic simulation (blue) and numerical solutions of the mode truncated equation (5) with different numbers of modes,  $M$ , are shown. The horizontal dashed lines represent the cumulative contributions for the all-infected case  $\Gamma_M^{\text{all}}$ . The vertical broken line at  $\beta_c = \mu/\lambda_1 \approx 4.6 \times 10^{-6} \text{ min}^{-1}$  is the transition point estimated from the largest eigenvalue of the averaged adjacency matrix  $\bar{A}$ .



**FIG. 3: Contribution of modes and verification of the final size equation.** a. Contribution for the all-infected case  $C_a^{\text{all}}$  versus eigenvalue  $\lambda_a$ . Although positively correlated,  $C_a^{\text{all}}$  is not a monotonically increasing function of  $\lambda_a$ . Arrows in this panel and panel b indicate the mode  $a = 8$ , which has smaller eigenvalue but larger contribution. The time evolution of this mode is depicted in Fig. 2c. b. Contribution  $C_a$  versus  $C_a^{\text{all}}$  obtained from the stochastic simulation. They are positively correlated, hence modes giving a high contribution for the all-infected case are important for the case of smaller prevalence. c. Cumulative contributions  $\Gamma_n^{\text{all}}$  (blue) and  $\Gamma_n$  (red) versus the number of modes  $n$ . An enlargement is shown in the inset. Contributions are heterogeneous, i.e. 10% of the total modes contribute approximately 90% in the all-infected case. d. Comparison of  $C_a$  obtained by the stochastic simulation and that obtained by the numerical solution of equation (5)  $\tilde{C}_a^{(100)}$  computed using 100 modes.



**FIG. 4: Identification of spatio-temporal hot spots and estimation of infection probability of test agents.** a. Spatial distribution of the risk factor  $\rho(\tau, \mathbf{x})$  of the proximity network at 00:00, 08:00, and 12:00, respectively. The upper and lower rows give the results of the stochastic simulation and those of the mode truncated equation with 100 modes, respectively. b. Schematic diagram to compute the risk factor  $\rho(\tau, \mathbf{x})$ . c. Infection probability of the  $p$ th test agent  $r_p(\infty)$  versus integral of the risk factor along the path of the agent  $\mathbf{x}_p(\tau)$ . Here,  $T = 1,440$  min is the normalisation constant. Dots are the numerical results, and the line is the theoretical prediction  $1 - \exp\left[-\frac{\beta}{\mu T} \int_0^T \rho(\tau, \mathbf{x}_p(\tau)) d\tau\right]$ . Details of the derivation of this equation is given in Supplementary Information SI5. d. The same as panel c, but the horizontal axis is the time integral of  $\tilde{\rho}^{(100)}$  obtained from the mode truncation with 100 modes.



# Supplementary information: Contribution of hidden modes to nonlinear epidemic dynamics in urban human proximity networks

## SCALING OF THE GIANT CLUSTER SIZE

The data size of the People flow data of  $N \approx 500,000$  is much smaller than the population of the corresponding Tokyo metropolitan area,  $N \approx 27,500,000$ . Therefore, it is natural to ask whether it makes sense to consider a model with a small number of agents. A clue to answering this question is to use scaling. We found that this system exhibits a percolation-like transition: the order parameter of the giant cluster component size ratio  $S(N, d)/N$  has a second-order phase transition (Fig. 5 left). Although it is unclear whether the percolation-like transition takes place for large  $N$  (because of the spatial resolution of the data for privacy reasons), we found that  $S(N, d)/N$  obeys the scaling law (Fig. 5 right). Far from the percolation transition point,  $S(N, d)/N \sim f(N^{0.605}d)$  holds, where  $f$  is a scaling function. If we assume that this scaling holds for the size of the real population, the parameter values  $N = 10,000$ ,  $d = 1,000$  m used in the simulations in the main text correspond to  $d \sim 10,000^{0.605} \times 1,000/27,500,000^{0.605} \approx 8.3$  m in the real population.

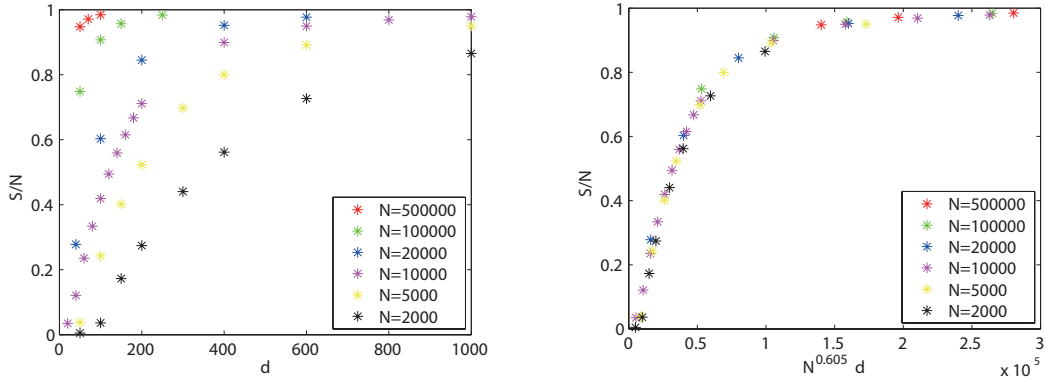


FIG. 5: Scaling of the giant cluster component size ratio for different  $N$ . The right-hand figure shows that this is a function of the scaled parameter  $N^{0.605}d$ .

## TIME-AVERAGED NETWORKS

We discuss an approximation to replace the time-dependent adjacency matrix with the time-averaged one, and determine the condition under which this approximation is valid.

Let us consider a general example where a dynamical variable  $\mathbf{X}(t)$  evolves via contact with other nodes as

$$\dot{\mathbf{X}}(t) = \beta A(t) \mathbf{X}(t), \quad (21)$$

where the parameter  $\beta$  represents the coupling strength and  $A(t)$  is the instantaneous adjacency matrix of a temporal network at time  $t$ . Equation (21) can be solved as

$$\mathbf{X}(t) = \exp_{\text{T}} \left[ \beta \int_0^t dt' A(t') \right] \mathbf{X}(0), \quad (22)$$

where  $\exp_{\text{T}}$  denotes the time-ordered exponential. If  $A(t)$  changes with the time interval  $\tau$ , equation (22) can be rewritten as

$$\begin{aligned} \mathbf{X}(L\tau) &= e^{\beta\tau A(t+(L-1)\tau)} e^{\beta\tau A(t+(L-2)\tau)} \dots e^{\beta\tau A(\tau)} e^{\beta\tau A(0)} \mathbf{X}(0) \\ &= (1 + \beta L\tau \bar{A}) \mathbf{X}(0) + O(\beta^2), \end{aligned} \quad (23)$$

where  $\bar{A} = \sum_{l=0}^{L-1} A(t + l\tau)/L$  is the time-averaged adjacency matrix. The above equation can be interpreted as follows. The terms of  $\beta^m$  in the power series expansion in equation (22) consist of time-ordered products of  $m$  adjacency matrices of the temporal network. These terms collect paths of length  $m$ , namely all possible indirect interactions occurring  $m$  times between two nodes at different times taking into account the time order of the interactions. If we consider the case of a small value of  $\beta$ , the time evolution can be approximated up to the first order. This corresponds to assuming that the interaction only takes place by direct contact between agents, and all multi-step interactions during the time interval  $T = L\tau$  are negligible. In such a case, the time-averaged network is sufficient to describe the dynamical processes in the temporal network. Note that the time-order information from one day does not appear in the first-order terms.

## NONLINEAR MODE TRUNCATED EQUATIONS

One can derive the mode coupling equation from the following differential equations (1) in the main text:

$$\begin{aligned}\frac{ds_j}{dt} &= -\beta s_j \sum_{k=1}^N \bar{A}_{jk} i_k \\ \frac{di_j}{dt} &= \beta s_j \sum_{k=1}^N \bar{A}_{jk} i_k - \mu i_j \\ \frac{dr_j}{dt} &= \mu i_j.\end{aligned}\tag{25}$$

Let us substitute equation (2) in the main text,

$$s_j(t) = \sum_{a=1}^N \hat{s}_a(t) \phi_j^{(a)}, \quad i_j(t) = \sum_{a=1}^N \hat{i}_a(t) \phi_j^{(a)}, \quad r_j(t) = \sum_{a=1}^N \hat{r}_a(t) \phi_j^{(a)},\tag{26}$$

into this equation. Taking into account the eigenvalue equation for the eigenvector  $\phi_j^{(a)}$ ,

$$\sum_{k=1}^N A_{jk} \phi_k^{(a)} = \lambda_a \phi_j^{(a)},\tag{27}$$

it is easy to verify that

$$\begin{aligned}\frac{d}{dt} \left[ \sum_{a'=1}^N \hat{s}_{a'} \phi_j^{(a')} \right] &= -\beta \left( \sum_{b=1}^N \hat{s}_b \phi_j^{(b)} \right) \left[ \sum_{k=1}^N \bar{A}_{jk}(t) \left( \sum_{c=1}^N \hat{i}_c \phi_k^{(c)} \right) \right] \\ &= -\beta \left( \sum_{b=1}^N \sum_{c=1}^N \lambda_c \phi_j^{(b)} \phi_j^{(c)} \hat{s}_b \hat{i}_c \right), \\ \frac{d}{dt} \left[ \sum_{a'=1}^N \hat{i}_{a'} \phi_j^{(a')} \right] &= \beta \left( \sum_{b=1}^N \hat{s}_b \phi_j^{(b)} \right) \left[ \sum_{k=1}^N \bar{A}_{jk}(t) \left( \sum_{c=1}^N \hat{i}_c \phi_k^{(c)} \right) \right] - \mu \sum_{a'=1}^N \hat{i}_{a'} \phi_j^{(a')} \\ &= \beta \left( \sum_{b=1}^N \sum_{c=1}^N \lambda_c \phi_j^{(b)} \phi_j^{(c)} \hat{s}_b \hat{i}_c \right) - \mu \sum_{a'=1}^N \hat{i}_{a'} \phi_j^{(a')}, \\ \frac{d}{dt} \left[ \sum_{a'=1}^N \hat{r}_{a'} \phi_j^{(a')} \right] &= \mu \sum_{a'=1}^N \hat{i}_{a'} \phi_j^{(a')}.\end{aligned}\tag{28}$$

Multiplying  $\phi_j^{(a)}$  on both sides of the above equations and taking the sum over  $j$  to use the orthonormal condition

$$\sum_{j=1}^N \phi_j^{(a)} \phi_j^{(b)} = \delta_{ab},\tag{29}$$

where  $\delta_{ac}$  is Kronecker's delta, we obtain

$$\begin{aligned}\frac{d\hat{s}_a}{dt} &= -\beta \left( \sum_{b=1}^N \sum_{c=1}^N \lambda_c G_{abc} \hat{s}_b \hat{i}_c \right), \\ \frac{d\hat{i}_a}{dt} &= \beta \left( \sum_{b=1}^N \sum_{c=1}^N \lambda_c G_{abc} \hat{s}_b \hat{i}_c \right) - \mu \hat{i}_a, \\ \frac{d\hat{r}_a}{dt} &= \mu \hat{i}_a,\end{aligned}\tag{30}$$

where  $G_{abc} = \sum_{j=1}^N \phi_j^{(a)} \phi_j^{(b)} \phi_j^{(c)}$ . Note that each element of  $G_{abc}$  consists of the product of the eigenvector components and is time-independent.

It is easy to extend this equation to the case where the static adjacency matrix  $\bar{A}$  is asymmetric. Let us define the left and right eigenvectors of  $\bar{A}$  for the eigenvalue  $\lambda_a$  as  $\psi_j^{(a)}$  and  $\phi_j^{(a)}$ , and set the normalisation condition  $\sum_{j=1}^N \psi_j^{(a)} \phi_j^{(b)} = \delta_{ab}$ . Then, equation (30) holds if we define  $G_{abc} = \sum_{j=1}^N \psi_j^{(a)} \phi_j^{(b)} \phi_j^{(c)}$ .

By considering  $M$  ( $\leq N$ ) modes and neglecting the other modes in equation (30), we obtain the transient dynamics of the epidemic from a small number of modes.

## DERIVATION OF THE LINEARISED EQUATION FROM THE MODE COUPLING EQUATION

Equations (30) contain mode coupling terms, but the modes are decoupled in the linearised equation. Let us take  $s_j(t) = 1 - \Delta s_j(t)$  and focus on the initial stage of the epidemic spreading. Assume that  $\Delta s_j(t)$  and  $i_j(t)$  are small, and approximate  $s_j i_j \approx i_j$  by neglecting the product  $\Delta s_j i_j$ . Since  $s_j$  appears only as the product  $s_j i_j$ , the linearised approximation allows us to put  $s_j \approx 1$ . Under this assumption, we can take

$$s_j = \sum_{b=1}^N \hat{s}_b \phi_j^{(b)} = 1.\tag{31}$$



Substituting this into the evolution equation for  $\hat{i}_a$  in equation (30), the linearised equation becomes

$$\frac{d\hat{i}_a}{dt} \approx \beta \left( \sum_{b=1}^N \sum_{c=1}^N \lambda_c G_{abc} \hat{s}_b \hat{i}_c \right) - \mu \hat{i}_a \quad (32)$$

$$= \beta \left( \sum_{j=1}^N \sum_{b=1}^N \sum_{c=1}^N \lambda_c \phi_j^{(a)} \phi_j^{(b)} \phi_j^{(c)} \hat{s}_b \hat{i}_c \right) - \mu \hat{i}_a. \quad (33)$$

By taking the summation over  $b$  using equation (31), we obtain

$$\frac{d\hat{i}_a}{dt} \approx \beta \left( \sum_{j=1}^N \sum_{c=1}^N \lambda_c \phi_j^{(a)} \phi_j^{(c)} \hat{i}_c \right) - \mu \hat{i}_a \quad (34)$$

$$= \beta \left( \sum_{c=1}^N \lambda_c \hat{i}_c \delta_{ac} \right) - \mu \hat{i}_a \quad (35)$$

$$= (\beta \lambda_a - \mu) \hat{i}_a. \quad (36)$$

In this way, we derive equation (3) in the main text.

## IDENTIFICATION OF HOT SPOTS AND ESTIMATION OF THE INFECTION PROBABILITY OF A TEST AGENT

We now discuss how to estimate the infection probability of a test agent. Let  $\mathbf{x}_j(t) = (x_j(t), y_j(t))^T$  be the position of agent  $j$  at time  $t$ . Since the infection probability depends on the distance between agents, we introduce the kernel function  $K(\mathbf{x})$ . In this study, we used the Heaviside step function  $K(\mathbf{x}) = \Theta(|\mathbf{x}| - D)$ , where  $\Theta(x) = 1$  for  $x \geq 0$  and  $\Theta(x) = 0$  otherwise. Then the element of the adjacency matrix  $A_{jk}(t)$  is given as  $A_{jk}(t) = K(\mathbf{x}_j(t) - \mathbf{x}_k(t))$ .

We define the spatio-temporal risk factor of infection as

$$\rho(\tau, \mathbf{x}) \equiv \sum_{k=1}^N r_k(\infty) K(\mathbf{x} - \mathbf{x}_k(\tau)). \quad (37)$$

where we take  $0 \leq \tau \leq T = 1440$  min, because we assume the mobility pattern is periodic over 24 h and  $\rho(\tau + T, \mathbf{x}) = \rho(\tau, \mathbf{x})$  holds.

We want to estimate the risk of infection of a test agent  $p$  who travels inside the urban area with an arbitrary trajectory and returns to the original location after time  $T$ . Equation

(1) in the main text can be solved as

$$s_j(t) = s_j(0)e^{-\beta\Psi_j(t)}, \quad (38)$$

where

$$\Psi_j(t) = \int_0^t \sum_{k=1}^N \bar{A}_{jk} i_k(t') dt' \quad (39)$$

$$= \frac{1}{\mu} \sum_{k=1}^N \bar{A}_{jk} [r_k(t) - r_k(0)]. \quad (40)$$

For  $t \rightarrow \infty$ ,  $i_j(\infty)$  should vanish, and the conservation of probability condition is given as  $r_j(\infty) = 1 - s_j(\infty)$ . Therefore, the infection probability of this agent is

$$r_j(\infty) = 1 - s_j(0) \exp \left\{ -\frac{\beta}{\mu} \sum_{k=1}^N \bar{A}_{jk} [r_k(\infty) - r_k(0)] \right\}. \quad (41)$$

This equation is transcendental and cannot be solved analytically. However, if  $N$  is sufficiently large and the contribution of the test agent to  $\rho(\tau, \mathbf{x})$  is negligible, we can approximate the infection probability of the test agent  $p$  using  $\rho(\tau, \mathbf{x})$ . Since the time-averaged adjacency matrix is given in terms of the risk factor as

$$\sum_{k=1}^N \bar{A}_{pk} r_k(\infty) = \frac{1}{T} \sum_{k=1}^N \int_0^T A_{pk}(\tau) r_k(\infty) d\tau \quad (42)$$

$$= \frac{1}{T} \int_0^T \rho(\tau, \mathbf{x}_p(\tau)) d\tau, \quad (43)$$

we obtain

$$r_p(\infty) = 1 - \exp \left[ -\frac{\beta}{\mu T} \int_0^T \rho(\tau, \mathbf{x}_p(\tau)) d\tau \right], \quad (44)$$

by taking  $s_j(0) = 1$  and  $r_k(0) = 0$ . Therefore, the final infection probability of a susceptible agent traveling along an arbitrary trajectory is given by the time integral of the risk factor along its trajectory.

If the mode truncation with  $M$  ( $\leq N$ ) modes

$$r_j(t) \approx \tilde{r}_j^{(M)}(t) \equiv \sum_{a=1}^M \hat{r}_{\sigma(a)}(t) \phi_j^{(\sigma(a))} \quad (45)$$

approximates  $r_j(t)$ , the risk factor is given by the superposition of  $M$  independent variables as

$$\rho(\tau, \mathbf{x}) \approx \tilde{\rho}^{(M)}(\tau, \mathbf{x}) \equiv \sum_{a=1}^M \sum_{j=1}^N \hat{r}_{\sigma(a)}(\infty) \phi_j^{(\sigma(a))} K(\mathbf{x} - \mathbf{x}_j(\tau)). \quad (46)$$

The infection probability is approximated with the mode truncation

$$r_p(\infty) \approx \tilde{r}_p(\infty) = 1 - \exp \left[ -\frac{\beta}{\mu T} \int_0^T \tilde{\rho}^{(M)}(\tau, \mathbf{x}_p(\tau)) d\tau \right], \quad (47)$$

by replacing  $\rho(\tau, \mathbf{x})$  with  $\tilde{\rho}^{(M)}(\tau, \mathbf{x})$  in equation (44).